5-15-2012

QUANTIFIYING THE BUSINESS IMPACT OF INFORMATION QUALITY - A RISK-BASED APPROACH

Valeria Klassen
Karlsruhe Institute of Technology

Alexander Borek
University of Cambridge

Ajith Parlikad

Robert Kern
Karlsruhe Institute of Technology

QUANTIFIYING THE BUSINESS IMPACT OF INFORMATION QUALITY - A RISK-BASED APPROACH

Klassen, Valeria, Karlsruhe Institute of Technology, Kaiserstraße 12, 76131 Karlsruhe, Germany, klassen.valeria@gmail.com

Borek, Alexander, University of Cambridge, The Old Schools, Trinity Lane, Cambridge CB2 1TN, UK, ab865@cam.ac.uk

Kern, Robert, Karlsruhe Institute of Technology, Kaiserstraße 12, 76131 Karlsruhe, Germany, robert.kern@ksri.uni-karlsruhe.de

Parlikad, Ajith Kumar, University of Cambridge, The Old Schools, Trinity Lane, Cambridge CB2 1TN, UK, aknp2@cam.ac.uk

Abstract

Effective decision making and business intelligence is highly depended on the quality of data and information available. Hence, data and information of poor quality can lead to poor decision making and is causing a variety of risks in every organization. Recent publications in the IS area have shown the link between information quality and risk from both theoretical and practical perspectives. This paper extends this work by providing a mathematical model based on extensive empirical data to model information risk. Moreover, we provide a practical example how the model can be operationalized to calculate the total risk of an information product based on our case study data. It is an important step towards a comprehensive business impact assessment of information quality, which would allow to build more sensible business cases for information quality improvements for managers.

Keywords: Business Impact of Information Quality, Risk Modeling, Information Risk Management.

1 Introduction to Information Risk Management

Information quality plays an important role for both business intelligence and knowledge management. Information risk is defined in this paper as the effect of uncertainty on objectives resulting from poor quality of information. Information can result from data in information systems or knowledge made explicit and communicated by humans in both structured and unstructured forms. In the IS discipline, data is typically defined as symbol or raw fact, whereas information is defined either as data that has been processed or as data plus meaning (Mingers 2006; Lewis 1991). Analogues to a traditional manufacturing system, where physical products are manufactured by using raw materials and processing them on an assembly line, information manufacturing is described as the process that transforms raw data into information products (Ballou et al. 1998; Wang 1998). Like physical products, the quality of information products should be measured and managed. Information quality (IQ) is defined from an user perspective as the fitness for use of information, in accordance to the IQ literature and is a multi-dimensional concept, e.g. accuracy, interpretability, completeness, security, up-to-dateness (Wang & Strong 1996). Recently, the link between information quality and risk has been established in IS literature (Borek et al. 2011; Borek et al. 2012), shown in *Figure 1*. When human, organizational and technological resources for information management are insufficient or are poorly coordinated (which would mean that the information management capabilities are poor), the quality of information suffers as a result (Ryu et al. 2006). Poor information quality can lead then to a variety of risks, which are called information risks in this paper, such as lowered customer and employee satisfaction, increased cost and difficulties in setting and executing strategy (Redman 1998) or even to major disasters like the explosion of space shuttle Challenger and shooting down of an Iranian Airbus by the USS Vincennes (Fisher & Kingma 2001). For instance, according to an in-depth study by GS1 UK, British companies in the retail sector are suffering from higher costs of at least 700 million Pounds and additionally loosing 300 million Pounds in revenue over 5 years due to poor quality information (GS1 UK 2009).

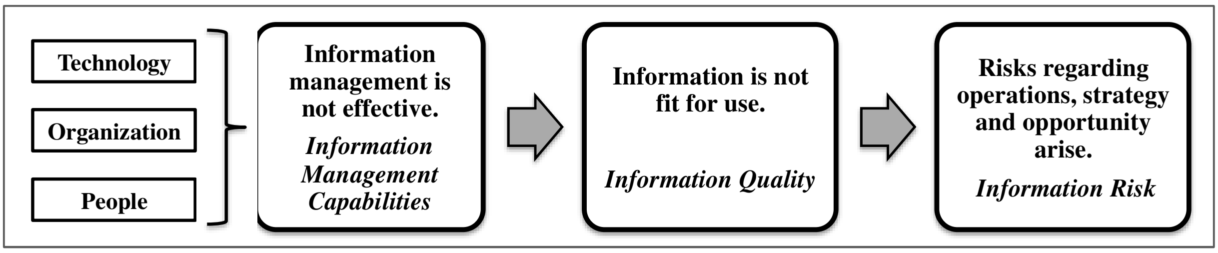

Figure 1. Interplay of information management, quality and risk; based on (Borek et al. 2011).

While there are quite established assessment methods for assessing the maturity of information management capabilities, e.g. (Baškarada 2009; Ryu et al. 2006; English 1999), and assessing information quality, e.g. (Pipino et al. 2002; Kaplan et al. 1998; Lee et al. 2002), there is a paucity of comprehensive methods for information risk assessment in the literature. A small number of practical approaches have been proposed by some consultants, e.g. (English 1999; McGilvray 2008; Loshin 2010). However, there has been no comprehensive model proposed for measuring the business impact of information quality so far and the probabilistic nature of information quality impacts has been widely ignored. Even et al. presented an utility driven approach to information quality (Even & Shankaranarayanan 2007), which provides a link between information quality and business impact, but suffers from the limitation that it is very hard to put utility functions into working practice. Some practitioners have emphasized the need to manage data and information quality as risk (Marinos 2004). Recently, a new approach has been presented that provides a full information risk management process to assess the risks that arise from poor information quality (Borek, Parlikad & Woodall 2011; Borek et al. 2011; Borek et al. 2011). The Total Information Risk Management (TIRM) process integrates best practices from the risk management and the information quality disciplines and offers step-by-step advice how to assess and treat information risks in organizations. A current weakness in this approach is, however, that there is no mathematical foundation provided for how to model and calculate information risks. This paper aims therefore at extending the TIRM process with a risk-based mathematical model. The model has been developed based on an extensive amount of empirical data collected by the authors in the industry. We collected the data using the TIRM process; the process and the case studies are described in more detail in (Borek, Parlikad & Woodall 2011). We have conducted five in-depth studies in the production and the energy sectors in the scope of engineering asset management, which is the management of complex physical assets over its whole lifecycle from planning and acquisition, usage, maintenance to disposal. In each of these studies, we found a large number of information risks that all follow a very similar pattern and can be described with our mathematical model. The remaining paper is structured as followed: First, we present the model and the calculations in the model. Then, we give an illustrative example based on our case study data and, hence, show how the model can be operationalized.

2 A Mathematical Model for Information Risk

2.1 Model Introduction

The information risk model consists of five different elements: an information product I^j, information quality problems $IQP(I^j)$, direct consequences and intermediate consequences of these problems $\mathbf{C^{BO}_{n(l),l}}$ and business objectives (BO). Figure 2 shows an example of the model for an information risk assessment scenario and is described in more detail in the following. Section 2.2 gives a formal definition of the elements of the model and section 2.3 discusses its assumptions. Moreover, we show how individual financial impacts (section 2.4) and the total risk of an information product (section 2.5) can be calculated. Finally, in section 2.6, we present how the financial impact on selected business objectives can be determined in our model. Risk is defined according to the definition of the International Standard Organization (ISO) as the "effect of uncertainty on objectives" (International Organization for Standardization 2009, p.9) and can be calculated as probability of a consequence multiplied with the impact of a consequence.

An information quality (IQ) problem arises when information is not fit for the specific purpose of a task and the outcome of the task is potentially influenced by this. One example that we encountered has been that data about the condition of production machines has been incorrect, which had an impact on decisions how to schedule maintenance activities. Root causes of an IQ problem are the technological, people or organizational factors that create an IQ problem (Lin et al. 2007) and, thus, alone or in combination, have the intrinsic potential to give rise to an information risk. In the case of the inaccurate machine condition data, the root causes were two-fold: (a) some sensors were not calibrated correctly and (b) engineering staff made mistakes when they entered the data manually into the system. An IQ problem can have one or more direct consequences and each direct consequence can have one or more intermediate consequences. A direct consequence is the immediate effect of an IQ problem, which has a likelihood attached. An intermediate consequence is a consequence of a consequence with a (conditional) likelihood attached. Note that intermediate consequences can cause further intermediate consequences. Some of the consequences have an impact on a business, which is measured as the impact on a defined business objective. Business objectives are strongly context-specific and are defined usually by senior management or the executive board. They can be financial goals, e.g. maximizing revenues, but may also include other aspects like product quality, delivery times, customer satisfaction and environmental objectives. The direct consequence in our example scenario was that decisions how to schedule preventive maintenance activities were sub-optimal. The intermediate consequence was in some cases that maintenance activities were executed unnecessary, which wasted money and which, thus, influenced the business objective "cost-effectiveness". Another intermediate consequence was that maintenance activities were not executed, although they would have been necessary, which could lead to machine failure and, in the worst case scenario, could cause the production to stop, which has an impact on "operational efficiency". Moreover, this could lead to a late delivery, which might impact the business objective "customer satisfaction".

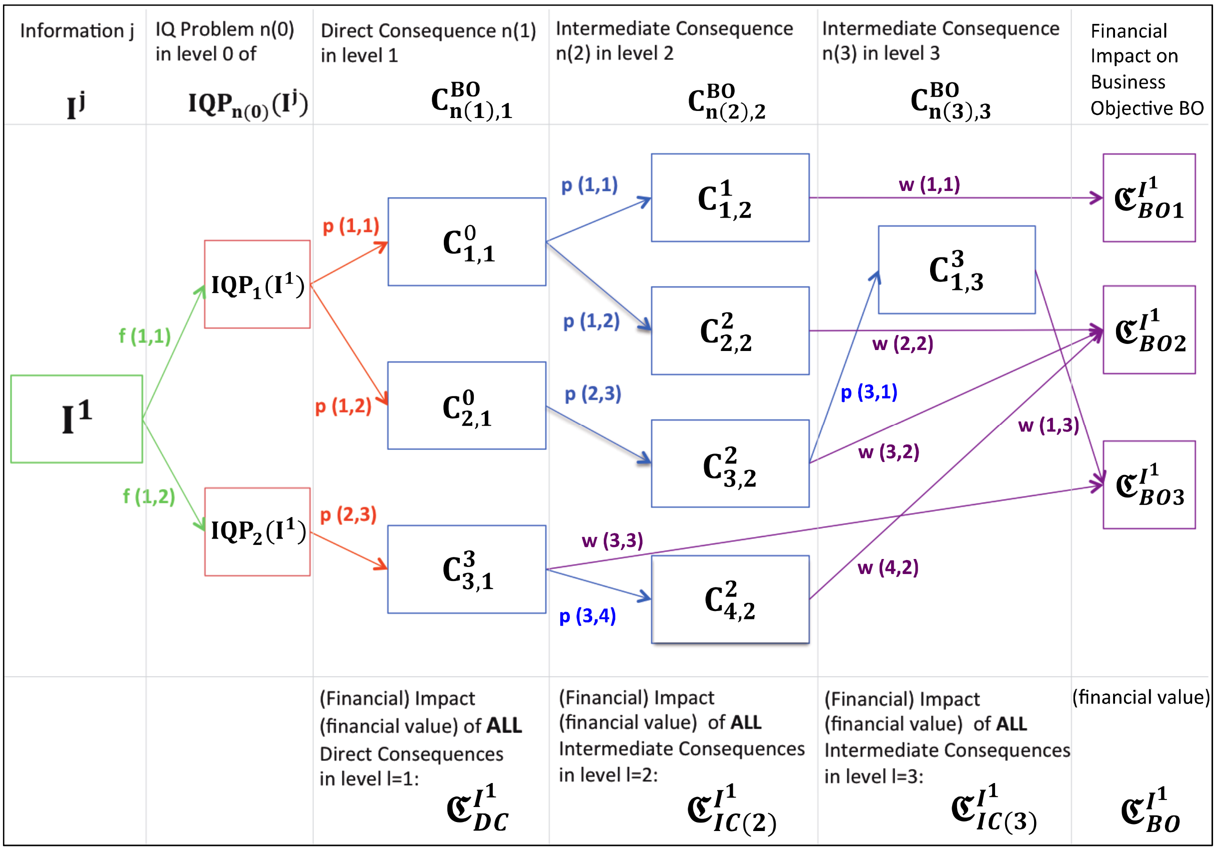

Figure 2. Overview of the Information Risk Model.

2.2 Basic Constructs

In the following, all variables used in this model, which are shown in Figure 2, are presented in detail and given a short practical interpretation.

$\mathbf{I^j}$: Information product j

Information products are the central point in our model and the total risk of poor information quality is determined individually for each information product. An information product can be manufactured manually and/or automatically by an information processing system using raw data. The organization has to define its information products depending on the context of usage. Depending on the area analyzed, an information product can be, for example, the information about which machines have been repaired provided by an operational information system, information about current market situation in form of a PDF report supplied by an external market research company, an overview of open order requests as shown in an Enterprise Resource Planning (ERP) system etc. An information product can be a high level aggregated information, e.g. information about all production machines that are maintained in the context of strategic asset purchasing decisions, or a collection of detailed specific data, e.g. about a single production machine as used for operational maintenance decisions. Since this IR assessment process might be very extensive, an indicator j is attached to each information product so that the list of the analyzed information in the business area can always be extended.

$\mathbf{IQP_{n(0)}(I^j)}$: IQ problem i of information product j

An IQ problem arises when an information product is not fit for the specific purpose of a task and this potentially influences the outcome of the task. By using this information product for a specific purpose, some problems might arise, called information quality problems (IQP). An indicator $n(0)$, number n in the list at level l=0, is attached to each IQP so that the analysis can stay dynamic and every time a new IQP can be added to the list. IQ problems are connected to a certain task the information product is used for and have a certain monthly frequency of occurrence $f(j, n(0))$ which can be derived, for example, by multiplying the frequency of use of the information product for the task per month with the probability that an IQ problem occurs.

IQ problems can be categorized using information quality dimensions that might be deficient, such as, accuracy, accessibility, completeness, information security, timeliness and understandability. For instance, machine repair data might be not up to date (timeliness) since the documents in the IT system environment are not updated automatically or the report about the current market situation might be not understandable (understandability) since the used financial abbreviations are not defined or the order request might be incomplete (completeness) since there exists no standard form in which all necessary order request details are considered and therefore also not accurate in some points if some characteristics can be interpreted in different ways.

$\mathbf{C}^{\mathbf{BO}}_{\mathbf{n(l),l}}$: Conditional expected financial value of the consequence n in level l

Due to each IQ problem one or more consequences can arise that we denote by $\mathbf{n(l)}$, these consequences are also random events with a conditional probability that might have a financial impact.

Note that an IQ problem has direct consequences (located in level 1) and intermediate consequences (located in level 2). The intermediate consequences can have further intermediate consequences (located in level 3…L). An IQ problem or a consequence can lead not just to one but also to multiple consequences. Each consequence of the accordant level is numerated by n(l).

Additionally, each consequence that has a financial impact for the organization influences a business objective, denoted by the indicator BO. Examples of business objectives are, for instance, “be highly profitable”, "exceeding our customers expectations" or "provide staff with safe working conditions". Note that each consequence can affect just one single business objective in the model, if it affects more than just one, it means that this consequence should be split into two or more consequences. In the case that a consequence does not affect any business objective, its value is set to 0, $BO \in \mathbb{N}$. The business objectives have to be defined by each organisation individually. The basic idea behind including business objectives in our model is that a manager might be especially interested to know which of the business objectives are affected negatively by poor information quality in order to align the priorities in IQ improvement to the current overall business priorities.

Assumptions in the Model

The model assumes that one information product j can have more that just one IQ problem $IQP_{n(0)}(I^j)$. These IQ problems are additionally assumed to be independent and can occur at the same time in an arbitrary constellation. If there are two IQ problems but which can arise just as a combination then we recommend to model them as one single IQ problem. The model further assumes that all consequences are independent. The reason for that assumption is more practice oriented. During several case studies, we discovered that, in general, it is already quite difficult for the experts to determine the probabilities of the occurrence of consequences. Therefore, it does not make really sense to consider dependency between these events as it adds unwanted complexity. It is well known from probability theory that the more complex a probability model gets, the more sensitive it is to inaccurate input which results in even more inaccurate output. Furthermore, we assume that a consequence, direct or intermediate, can have one or more following consequences that are also independent and can arise at the same time in an arbitrary constellation. Finally, one of the learning points from our case studies was that it would not be feasible to get data input to a continuous model, which made us choose a discrete approach.

Financial impacts of a consequence can be classified regarding a business objective they are affecting. Understanding which information quality problems have an impact on which business objectives allows a prioritization of information risk treatment initiatives based on the current priorities of the executive management board. A detailed classification of business impacts is provided by Loshin (Loshin 2010) which however needs to be adapted to the context of the organization. For each business objective a metric needs to be defined by the organization. A metric can be financial or non-financial and use a quantitative or /and a qualitative scale. Notice that especially direct consequences often do not cause a financial impact and impact on business objectives because in many cases they represent a decision, which is influenced by the regarded information quality problem. It yet often causes an intermediate consequence, which then has a financial impact and an observable impact on one of the business objectives. The consequences can be, for example, “the maintenance plan cannot be designed correctly due to outdated machine repair data” ($C_{1,1}$), followed by another consequences: “machine failure” ($C_{2,1}^{01}$) which is then followed by multiple other consequences: “machine breakdown” ($C_{3,1}^{01}$), “damaged products” ($C_{3,2}^{04}$), “decelerated production process” ($C_{3,3}^{03}$) and so on. The goal of this information risk model is to calculate the financial impact of poor information quality in the business area that is analyzed and to understand the impact on business objectives. Since consequences create financial costs, the notation $C_{l,n(l)}^{BO}$ includes not just the description of the consequence but also the financial impact, which will always be addressed using this variable in formulas. In the model, frequencies are used between the information product j and the accordant IQ problems. Probabilities are used between the consequences. All probabilities are considered to be conditional probabilities. They can be described by a relative frequency of occurrence of the pointed instances; this probability is denoted by $p_{[\cdot,\cdot]}$. The complementary probability is represented by $\bar{p}_{[\cdot,\cdot]}$ and has the meaning that the pointed object will not occur, therefore: $p_{[\cdot,\cdot]} + \bar{p}_{[\cdot,\cdot]} = 1$. Since the complementary probability is not necessary for the calculation of information risk, it does not appear in the model.

2.3 Determining the Individual Financial Impact $C_{n(l),l}^{BO}$

The conditional, expected financial value $C_{n(l),l}^{BO}$ is determined using three different scenarios, “low” (L_{fi}) represents a typical scenario that has a rather low impact, “medium” (M_{fi}) a scenario that has an average impact and “high” (H_{fi}) that has a quite high impact, as shown in Figure 3. Each scenario has a probability attached: q_{LOW} / q_{MEDIUM}/ q_{HIGH}, which is the probability that the consequence n(l) at level l will cause a low/medium/high financial impact. $C_{n(l),l}^{BO}$ is then the conditional expected financial impact that the consequence n(l) at level l will cause when it occurs, which is:

$$\mathbf{C_{n(l),l}^{BO} = q_{LOW} * L_{fi} + q_{MEDIUM} * M_{fi} + q_{HIGH} * H_{fi}}$$

$$0 \le q_{LOW} \le 1$$
$$0 \le q_{MEDIUM} \le 1$$
$$0 \le q_{HIGH} \le 1$$
$$q_{LOW} + q_{MEDIUM} + q_{HIGH} = 1$$

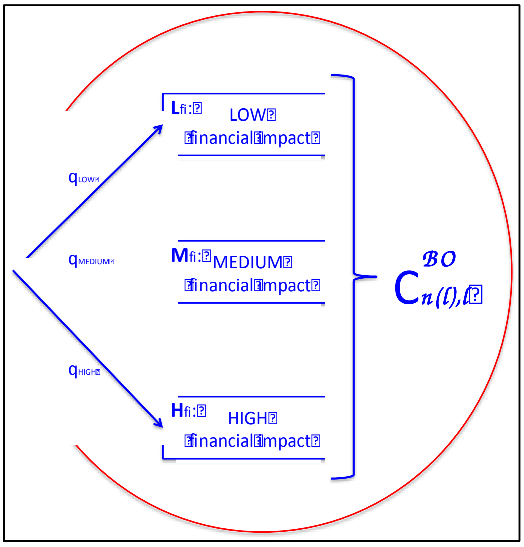

Figure 3: Calculation of conditional expected financial value of consequence n in level l, $C_{n(l),l}^{BO}$.

2.4 Calculating the Total Expected Risk

Summing up all consequences together, direct and intermediate at all levels gives us the total financial impact of poor information quality, caused by information product j, which we call the total expected risk $\boldsymbol{R}(\boldsymbol{I^j})$:

$$\boldsymbol{R}(\boldsymbol{I^j}) = \boldsymbol{\mathfrak{C}^{I^j}_{DC}} + \sum_{\boldsymbol{l=2}}^{\boldsymbol{L}} \boldsymbol{\mathfrak{C}^{I^j}_{IC(l)}}$$

Direct consequences are caused directly from the accordant IQ problems and are defined in the first level, l=1. The variable $\boldsymbol{\mathfrak{C}^{I^j}_{DC}}$ represents the financial impact over all direct consequences caused by poor information quality of information product j.

$$\boldsymbol{\mathfrak{C}^{I^j}_{DC}} = \sum_{n(0)=1}^{N(0)} \sum_{n(1)=1}^{N(1)} f_{I^j,n(0)} \cdot p_{[n(0),0],[n(1),1]} \cdot C^{BO}_{n(1),1}$$

$$f_{I^j,n(0)} \in \mathbb{N}_{>0} \qquad 0 \leq p_{[\cdot,\cdot]} \leq 1, \qquad N(\cdot) \in \mathbb{N}, \quad C^{BO}_{n(1),1} \geq 0$$

The variable $N(0)$ represents the total number of IQ problems of the information product j and $N(l)$ represents the total number of consequences in level l. The monthly frequency of IQ problem $n(0)$ is denoted by $f_{I^j,n(0)}$. The probability that the direct consequence $C^{BO}_{n(1),1}$ will arise due to the IQ problem $n(0)$ is denoted by $p_{[n(0),0],[n(1),1]}$.

Since there can be multiple levels of intermediate consequences, a general formula is given to calculate the financial impact of all intermediate consequences at level l, $l \in \{2, \dots, L\}$, and L denotes the highest level of the intermediate consequences in the present model. Therefore the variable $\boldsymbol{\mathfrak{C}^{I^j}_{IC(l)}}$ represents the sum of the financial impact of all intermediate consequences at level l, caused by poor quality of information product j.

$$\boldsymbol{\mathfrak{C}^{I^j}_{IC(l)}} = \sum_{n(0)=1}^{N(0)} \dots \sum_{n(l-1)=1}^{N(l-1)} \sum_{n(l)=1}^{N(l)} \left(f_{I^j,n(0)} \cdot \prod_{t=1}^{l} p_{[n(t-1),t-1],[n(t),t]} \cdot C^{BO}_{n(l),l} \right)$$

$$f_{I^j,n(0)} \in \mathbb{N}_{>0} \qquad 0 \leq p_{[\cdot,\cdot]} \leq 1, \qquad N(l) \in \mathbb{N}, \qquad l \in \{2, \dots, L\}, \qquad C^{BO}_{n(l),l} \geq 0, \qquad t \in \mathbb{N}$$

2.5 Determining the Financial Impact on Specific Business Objectives

Additionally, the overall expected financial impact on a specific business objective BO can be calculated by summing up the expected impacts of consequences affecting this certain BO. Depending on the BO, these consequences can represent a financial value or an intangible risk. In detail, over all levels, the impacts of consequences affecting this certain BO need to be summed up, considering for each of them all the conditional probabilities they depend on. The variable $w_{[n(l),l],BO}$ represents the connection between the consequence $n(l)$ in level l and the certain BO. In the case that the consequence $C^{BO}_{n(l),l}$ has an impact on the business objective BO, their connection is set to one, $w_{[\cdot],BO} = 1$, otherwise it is zero, $w_{[\cdot],BO} = 0$.

$$\boldsymbol{\mathfrak{C}^{I^j}_{BO}} = \boldsymbol{\mathfrak{C}^{I^j}_{DC,BO}} + \sum_{\boldsymbol{l=2}}^{\boldsymbol{L}} \boldsymbol{\mathfrak{C}^{I^j}_{IC(l),BO}}$$

The financial impact on a certain business objective BO created by all direct consequences arising due to poor IQ of information product j is determined as follows:

$$\mathfrak{C}^{I^j}_{\mathrm{DC},\boldsymbol{BO}} = \sum_{n(0)=1}^{N(0)} \sum_{n(1)=1}^{N(1)} f_{I^j,n(0)} \cdot p_{[n(0),0],[n(1),1]} \cdot w_{[n(1),1],BO} \cdot C^{BO}_{n(1),1}$$

$$f_{I^j,n(0)} \in \mathbb{N}_{>0} \quad 0 \le p_{[\cdot,\cdot]} \le 1, \quad N(\cdot) \in \mathbb{N}, \quad C^{BO}_{n(1),1} \ge 0, \quad w_{[\cdot],BO} \in \{0,1\}$$

Furthermore, the financial impact of all intermediate consequences for each single level l influencing the certain business objective BO needs to be calculated by using this formula:

$$\mathfrak{C}^{I^j}_{\boldsymbol{IC(l)},\mathbf{BO}} = \sum_{n(0)=1}^{N(0)} \cdots \sum_{n(l-1)=1}^{N(l-1)} \sum_{n(l)=1}^{N(l)} \left(f_{I^j,n(0)} \cdot \prod_{t=1}^{l} p_{[n(t-1),t-1],[n(t),t]} \cdot w_{[n(l),l],BO} \cdot C^{BO}_{n(l),l} \right)$$

$$f_{I^j,n(0)} \in \mathbb{N}_{>0} \quad 0 \le p_{[\cdot,\cdot]} \le 1, \quad N(\cdot) \in \mathbb{N}, \quad C^{BO}_{n(l),l} \ge 0, \quad w_{[\cdot],BO} \in \{0,1\}, \quad l \in \{2,\ldots,L\}$$

If a company has current or longer-term priorities regarding specific business objectives, e.g. a company needs to improve its product quality, it is interesting to look at the financial impact on the specific business objective, here “Product Quality”. The calculated financial impact would show to which level poor IQ of information product j negatively influences this BO.

3 Example

In this section, we give a representative example based on one of our case studies to illustrate how the mathematical model can be operationalized, visualized in Figure 2. This particular scenario is in the context of an energy provider that needs to maintain large, complex physical assets that are distributed geographically. Numeric values have been modified out of confidentiality reasons, keeping the reality of the scenario intact.

3.1 Data Input into Model

Information product $\mathbf{I^j}$

$\mathbf{I^1}$	**Geographical information about assets**

Information quality problems, $\mathbf{IQP_{n(0)}(I^j)}$

$\mathbf{IQP_1(I^1)}$:	**Accuracy and Completeness**: Information on the plans, cable data are sometimes inaccurate. Staff does not always complete the forms. Especially data from many years ago is missing about the assets, it does not always tell you the type of asset and age of asset.
$\mathbf{IQP_2(I^1)}$	**Accessibility**: Data is not accessible in the field during maintenance, site visits, faults etc. Equipment is not available to do so. This is a permanent problem.

Monthly frequency of occurrence of the accordant IQ problem, $\boldsymbol{f_{I^1,n(0)}}$

$\mathbf{f_{I^1,1}}$	**1050 times per month**
$\mathbf{f_{I^1,2}}$	**240 times per month**

Business objectives that are impacted

$\boldsymbol{BO}\,\mathbf{1}$	Operational Cost Efficiency
$\boldsymbol{BO}\,\mathbf{2}$	Health and Safety
$\boldsymbol{BO}\,\mathbf{3}$	Customer Satisfaction

Direct consequences: $\boldsymbol{C_{n(1),1}^{BO}}$

$\boldsymbol{C_{1,1}^{0}}$	Poor strategic and tactical maintenance decisions as it is based on incomplete and inaccurate geographical data
$\boldsymbol{C_{2,1}^{0}}$	Poor operational maintenance decisions as it is based on incomplete and inaccurate geographical data
$\boldsymbol{C_{3,1}^{3}}$	Engineers cannot make maintenance decisions in the field due to inaccessibility of geographical data

(Note that BO3 in $\boldsymbol{C_{3,1}^{3}}$ indicates that this direct consequence has a direct impact on business objective 3, which is associated with a financial impact.)

Frequency of the IQ problem $\mathbf{IQP_{n(0)}(I^1)}$ will cause the Direct Consequence $\boldsymbol{C_{n(1),1}^{BO}}$, $\boldsymbol{p}_{[\boldsymbol{n}(\mathbf{0}),\mathbf{0}],[\boldsymbol{n}(\mathbf{1}),\mathbf{1}]}$

$\boldsymbol{p}_{[\mathbf{1},\mathbf{0}],[\mathbf{1},\mathbf{1}]}$	**20%** (probability that IQ problem 1 “Accuracy and Completeness” leads to direct consequence 1 “Poor strategic and tactical maintenance decisions”)
$\boldsymbol{p}_{[\mathbf{1},\mathbf{0}],[\mathbf{2},\mathbf{1}]}$	**35%** (probability that IQ problem 1 “Accuracy and Completeness” leads to direct consequence 2 “Poor operational maintenance decisions”)
$\boldsymbol{p}_{[\mathbf{2},\mathbf{0}],[\mathbf{3},\mathbf{1}]}$	**80%** (probability that IQ problem 2 “Accessibility” leads to direct consequence 3 “Engineers cannot make maintenance decisions in the field”)

All other possible connections between the presented IQ problems and the shown direct consequences have the probability equal to zero.

Intermediate consequences at level 2: $\mathbf{C_{n(2),2}^{BO}}$

$\boldsymbol{C_{1,2}^{1}}$	Higher maintenance and repair costs due to suboptimal maintenance plans
$\boldsymbol{C_{2,2}^{2}}$	People could get injured
$\boldsymbol{C_{3,2}^{2}}$	Repairs take longer
$\boldsymbol{C_{4,2}^{2}}$	Site revisits by engineers are necessary

Probability that direct consequence $\boldsymbol{C_{n(1),1}^{BO}}$ causes intermediate consequence at level 2, $\boldsymbol{p}_{[\boldsymbol{n}(\mathbf{1}),\mathbf{1}],[\boldsymbol{n}(\mathbf{2}),\mathbf{2}]}$

$\boldsymbol{p}_{[\mathbf{1},\mathbf{1}],[\mathbf{1},\mathbf{2}]}$	**30%** (probability that direct consequence 1 “Poor strategic and tactical maintenance decisions” at level 1 leads to intermediate consequence 1 “Higher maintenance and repair costs due to suboptimal maintenance plans” at level 2)
$\boldsymbol{p}_{[\mathbf{1},\mathbf{1}],[\mathbf{2},\mathbf{2}]}$	**10%** (probability that direct consequence 1 “Poor strategic and tactical

maintenance decisions” at level 1 leads to intermediate consequence 2 “People could get injured” at level 2)

$p_{[2,1],[3,2]}$ **25%** (probability that direct consequence 2 “Poor operational maintenance decisions” at level 1 leads to intermediate consequence 3 “Repairs take longer” at level 2)

$p_{[3,1],[4,2]}$ **40%** (probability that direct consequence 3 “Engineers cannot make maintenance decisions in the field” at level 1 leads to intermediate consequence 4 “Site revisits by engineers are necessary” at level 2)

All other possible connections between the presented direct consequences and the shown intermediate consequences have the probability equal to zero.

Intermediate Consequences at level 3: $\mathbf{C}^{\mathbf{BO}}_{\mathbf{n(3),3}}$

$\mathbf{C}^{3}_{1,3}$ Customer minutes are lost (a customer is not connected to the electricity grid for a certain amount of time)

Probability that the Intermediate Consequence $C^{BO}_{n(2),2}$ at level 2 will cause the Intermediate Consequence at level 3, $p_{[n(2),2],[n(3),3]}$

$p_{[3,2],[1,3]}$ **10%** (probability that direct consequence 3 “Repairs take longer” at level 2 leads to intermediate consequence 1 “Customer minutes are lost” at level 3)

All other possible connections between the presented intermediate consequences at level 2 and the shown intermediate consequences at level 3 have the probability equal to zero.

3.2 Output: Financial Values

Table 1 shows how the conditional, expected financial individual values are calculated using the input data from the last section.

Conditional, expected individual financial values (on monthly basis) $C^{BO}_{n(l),l} = = q_{LOW} * L_{fi} + q_{MEDIUM} * M_{fi} + q_{HIGH} * H_{fi}$		
$C^{BO3}_{3,1}$=	**55% * 5 € + 40% * 30 € + 5% * 500 €**	= **40** € Engineers cannot make maintenance decisions in the field. ***Impact on BO: BO3 “Customer Satisfaction”.***
$C^{BO1}_{1,2}$=	**30% * 0 € + 67% * 100 € + 3% * 50000 €**	= **1567 €** Higher maintenance and repair costs. ***Impact on business objective: BO1 “Operational Cost Efficiency”.***
$C^{BO2}_{2,2}$=	**80% * 0 € + 19% * 1000 € + 1% * 30000€**	= **490 €** People get injured or killed. ***Impact on business objective: BO2 “Health & Safety.***
$C^{BO1}_{3,2}$=	**40% * 40€ + 50% * 200€ + 10% * 5000€**	= **616 €** Repairs take longer. ***Impact on business objective: BO1 “Operational Cost Efficiency”.***
$C^{BO1}_{4,2}$=	**10% * 30€ + 80% * 90€ + 10% * 200€**	= **95 €** Site revisits by engineers are necessary. ***Impact on business objective: BO1 “Operational Cost Efficiency”.***
$C^{BO3}_{1,3}$=	**30% * 5 € + 60% * 20 € + 10% * 500 €**	= **63** € Customer minutes are lost. ***Affected business objective: BO3 “Customer Satisfaction”.***

Table 1. Conditional, expected individual financial values

Table 2 illustrates how the Total Expected Risk can be determined using the input data in our case study scenario.

Calculation of the Total Expected Risk (on monthly basis) $\boldsymbol{R(I^1)} = \mathfrak{C}^{I^j}_{DC} + \sum_{l=2}^{3} \mathfrak{C}^{I^1}_{IC(l)} = 7{,}680\ € + 172{,}902€ + 578\ € = \mathbf{181{,}160\ €\ monthly}$	
$\mathfrak{C}^{I^1}_{DC} =$	$240 \text{ times per month} * 80\% * 40\ € = \mathbf{7{,}680\ €\ per\ month}$
$\mathfrak{C}^{I^1}_{IC(2)} =$	$1050 \text{ times per month} * \big(20\% * (30\% * 1567\ € + 10\% * 490\ €)\big) + 35\% * (25\% * 616\ €) + 240 \text{ times per month} * 80\% * (40\% * 95\ €) = \mathbf{172{,}902€\ per\ month}$
$\mathfrak{C}^{I^1}_{IC(3)} =$	$1050 \text{ times per month} * 35\% * 25\% * 10\% * 63\ € = \mathbf{578\ €\ per\ month}$

Table 2. Calculation of the Total Expected Risk

A final remark at the end: As with all models, the quality of the output is highly depended on the quality of the input data, known as the "garbage in – garbage out" principle. This paper is, however, primarily concerned with providing a mathematical foundation to calculate the financial impact of information risks. Strategies to obtain a better quality of the input data in information risk assessment for the model, such as using numeric ranges for better estimation of the parameters, including objective numeric values and cross-checking of results, are discussed in (Borek, Parlikad & Woodall 2011).

4 Conclusion

Due to the massive amount of data available and the rising capabilities and complexity of information systems and organizations, information quality creates increasingly risks in every organization on a operational and strategic level and organizations who want to create a competitive advantage have to understand how to manage these risks effectively (Redman 2008). A prerequisite of effective information risk management is the assessment of these risks. This paper has presented a risk-based approach to model and calculate the financial impact of information quality problems in organizations, which, so far, has been a very unexplored area that is yet of high relevance for data managers and information quality practitioners. Our next step is to integrate the mathematical model in a mind-mapping based software tool that we are currently developing to assist in the assessment of information risks. The software tool will be then tested in several industrial companies.

References

Ballou, D. et al., 1998. Modeling information manufacturing systems to determine information product quality. *Management Science*, 44(4), pp.462–484.

Baškarada, S., 2009. *IQM-CMM: Information Quality Management Capability Maturity Model*, Vieweg +Teubner.

Borek, A. et al., 2011. An information oriented framework for relating IS/IT resources and business value. In *Proceedings of the International Conference on Enterprise Information Systems (ICEIS)*. International Conference on Enterprise Information Systems (ICEIS). Beijing, China.

Borek, A. et al., 2012. IS/IT Resources and Business Value: Operationalisation of an Information Oriented Framework. In *Enterprise Information Systems - 13th International Conference, ICEIS 2011, Beijing, China, June 8-11, 2011, Revised Selected Papers*. Lecture Notes in Business Information Processing. Springer.

Borek, A., Parlikad, A.K. & Woodall, P., 2011. Towards A Process For Total Information Risk Management (TIRM). In *Proceedings of the 16th International Conference on Information Quality (ICIQ 2011)*. Adelaide, Australia.

Borek, A., Woodall, P. & Parlikad, A.K., 2011. A Risk Management Approach to Improving Information Quality for Operational and Strategic Management. In *Proceedings of the 18th EUROMA Conference: Exploring Interfaces*. EurOMA Conference 2011. Cambridge, UK.

Borek, A., Woodall, P. & Parlikad, A.K., 2011. Managing Information Risks in Asset Management - Experiences from an In-depth Case Study in the Utility Industry. In *Proceedings of the IET and IAM Asset Management conference 2011*. London, UK.

English, L.P., 1999. *Improving data warehouse and business information quality: methods for reducing costs and increasing profits*, John Wiley & Sons.

Even, A. & Shankaranarayanan, G., 2007. Utility-driven assessment of data quality. *SIGMIS Database*, 38(2), pp.75–93.

Fisher, C.W. & Kingma, B.R., 2001. Criticality of data quality as exemplified in two disasters. *Information & Management*, 39(2), pp.109–116.

GS1 UK, 2009. Data Crunch Report - The Impact of Bad Data on Profits and Consumer Service in the UK Grocery Industry. Available at: http://www.gs1uk.org/resources/help_support/Pages/White-Papers.aspx.

International Organization for Standardization, 2009. ISO Guide 73:2009 - Risk management -- Vocabulary. Available at: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=44651 [Accessed April 25, 2011].

Kaplan, D. et al., 1998. Assessing data quality in accounting information systems. *Commun. ACM*, 41(2), pp.72–78.

Lee, Y.W. et al., 2002. AIMQ: a methodology for information quality assessment. *Information & Management*, 40(2), pp.133–146.

Lewis, P.J., 1991. The decision making basis for information systems: the contribution of Vickers' concept of appreciation to a soft systems perspective. *European Journal of Information Systems*, 1(1), pp.33–44.

Lin, S. et al., 2007. Developing a data quality framework for asset management in engineering organisations. *International Journal of Information Quality*, 1(1), pp.100–126.

Loshin, D., 2010. *The Practitioner's Guide to Data Quality Improvement*, Morgan Kaufmann.

Marinos, G., 2004. Data Quality: A Risk-Based Approach. *Information Management*. Available at: http://www.information-management.com/issues/20040801/1007208-1.html.

McGilvray, D., 2008. *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*, Morgan Kaufmann.

Mingers, J., 2006. *Realising systems thinking: knowledge and action in management science*, Springer.

Pipino, L.L., Lee, Y.W. & Wang, R.Y., 2002. Data quality assessment. *Communications of the ACM*, 45(4), pp.211–218.

Redman, T.C., 2008. *Data Driven: Profiting from Your Most Important Business Asset* illustrated ed., Harvard Business School Press.

Redman, T.C., 1998. The impact of poor data quality on the typical enterprise. *Communications of the ACM*, 41(2), pp.79–82.

Ryu, K.S., Park, J.S. & Park, J.H., 2006. A data quality management maturity model. *ETRI journal*, 28(2), pp.191–204.

Wang, R.Y., 1998. A product perspective on total data quality management. *Communications of the ACM*, 41(2), pp.58–65.

Wang, R.Y. & Strong, D.M., 1996. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4), p.33.